

STATISTICS FORMULA SHEET

CA Nishant Kumar

Stay updated with the latest FREE lectures: <https://www.youtube.com/@ekagrataca>

CA NISHANT KUMAR

Chapter 13 – Statistical Description of Data

1. Frequency Density of a Class Interval = $\frac{\text{Class Frequency}}{\text{Class Length}}$
2. Relative Frequency of a Class Interval = $\frac{\text{Class Frequency}}{\text{Total Frequency}}$
3. Percentage Frequency of a Class Interval = $\frac{\text{Class Frequency}}{\text{Total Frequency}} \times 100$

Chapter 14 – Measures of Central Tendency and Dispersion

Topic 1 – Arithmetic Mean

1. Mean of Individual Series $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

2. Mean of Discrete/Continuous Series $\bar{x} = \frac{\sum fx}{\sum f}$

(In continuous series, the mid-point of the class interval is taken to be x)

3. If two variables x and y are related as $y = a + bx$, and $\bar{x}$ is known, then $\bar{y} = a + b\bar{x}$.

4. Combined Mean $\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

Topic 2 – Geometric Mean

1. GM of Individual Series = n^{th} root of the product of n numbers

2. GM of Discrete/Continuous Series = $\left(x_1^{f_1} \times x_2^{f_2} \times \dots \times x_n^{f_n}\right)^{\frac{1}{N}}$

(In continuous series, the mid-point of the class interval is taken to be x)

3. If $z = xy$, then $GM \text{ of } z = GM \text{ of } x \times GM \text{ of } y$

4. If $z = \frac{x}{y}$, then $GM \text{ of } z = \frac{GM \text{ of } x}{GM \text{ of } y}$

5. For a set of r observations, $\log G = \frac{1}{r} \sum \log x$

Topic 3 – Harmonic Mean

1. HM of individual series $HM = \frac{n}{\sum\left(\frac{1}{x}\right)}$

2. HM of Discrete/Continuous Series $HM = \frac{N}{\sum\left(\frac{f}{x}\right)}$

(In continuous series, the mid-point of the class interval is taken to be x)

3. Combined $HM = \frac{n_1 + n_2}{\frac{n_1}{H_1} + \frac{n_2}{H_2}}$

4. The harmonic mean of $1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}$ is given by: $\frac{2}{n+1}$

5. Average Speed = $\frac{2xy}{x+y}$

6. Relationship between AM, GM, HM for constant observations: $AM = GM = HM$

7. Relationship between AM, GM, HM for distinct observations: $AM > GM > HM$

8. For positive observations, $GM^2 = AM \times HM$

Topic 4 – Median

1. For odd number of items, Rank of Median of Individual Series: $\frac{n+1}{2}$
2. For even number of items, Median is given by the average of $\left(\frac{n}{2}\right)^{th}$ term and $\left(\frac{n}{2}+1\right)^{th}$ term.
3. Rank for Discrete Series: $\frac{N+1}{2}$
4. Rank for Continuous Series: $\frac{N}{2}$

5. Median of Continuous Series: $l + \frac{\text{Rank} - c}{f} \times i$

6. If two variables x and y are related as $y = a + bx$, and x_{me} is known, then
 $y_{me} = a + bx_{me}$.

Topic 5 – Quartile

1. Rank of Q_1 of Individual Series: $\frac{n+1}{4}$
2. Rank of Q_1 of Discrete Series: $\frac{N+1}{4}$
3. Rank of Q_1 of Continuous Series: $\frac{N}{4}$
4. Rank of Q_2 & Q_3 : $2 \times \text{Rank of } Q_1$, or $3 \times \text{Rank of } Q_1$
5. Quartile of Continuous Series: $l + \frac{\text{Rank} - c}{f} \times i$

Topic 6 – Decile

1. Rank of D_1 of Individual Series: $\frac{n+1}{10}$
2. Rank of D_1 of Discrete Series: $\frac{N+1}{10}$
3. Rank of D_1 of Continuous Series: $\frac{N}{10}$
4. Rank of D_2 & D_3 : $2 \times \text{Rank of } D_1$, or $3 \times \text{Rank of } D_1$
5. Decile of Continuous Series: $l + \frac{\text{Rank} - c}{f} \times i$

Topic 7 – Percentile

1. Rank of P_1 of Individual Series: $\frac{n+1}{100}$

2. Rank of P_1 of Discrete Series: $\frac{N+1}{100}$

3. Rank of P_1 of Continuous Series: $\frac{N}{100}$

4. Rank of P_2 & P_3 : $2 \times \text{Rank of } P_1$, or $3 \times \text{Rank of } P_1$

5. Percentile of Continuous Series: $l + \frac{\text{Rank} - c}{f} \times i$

Topic 8 – Mode

1. Mode of Continuous Series = $l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$
2. If two variables x and y are related as $y = a + bx$, and x_{mo} is known, then $y_{mo} = a + bx_{mo}$.
3. Relationship between Mean, Median, Mode
 - a. For Symmetric Data: Mean = Median = Mode
 - b. For Skew Symmetric Data: Mode = 3Median – 2Mean
 - c. For Positively Skewed Data: Mean > Median > Mode
 - d. For Negatively Skewed Data: Mean < Median < Mode

Topic 9 – Range

1. Range of Individual Series = Largest Observation – Smallest Observation
2. Coefficient of Range = $\frac{\text{Largest Observation} - \text{Smallest Observation}}{\text{Largest Observation} + \text{Smallest Observation}}$
3. Range of Continuous Series = Upper Most Class Boundary – Lower Most Class Boundary
4. Coefficient of Range = $\frac{UMCB - LMCB}{UMCB + LMCB}$
5. If two variables x and y are related as $y = a + bx$, and R_x is known, then $R_y = |b| \times R_x$.

Topic 10 – Mean Deviation

1. Mean Deviation of Individual Series = $\frac{\sum |x - A|}{n}$

2. Mean Deviation of Discrete/Continuous Series = $\frac{\sum f |x - A|}{N}$

3. Coefficient of Mean Deviation $\frac{\text{Mean Deviation About } A}{A} \times 100$

4. If two variables x and y are related as $y = a + bx$, and MD_x is known, then
 $MD_y = |b| \times MD_x$.

Topic 11 – Standard Deviation

1. Standard Deviation of Individual Series = $\sqrt{\frac{\sum (x - \bar{x})^2}{n}}$

2. Standard Deviation of Individual Series = $\sqrt{\frac{\sum x^2}{n} - (\bar{x})^2}$

3. Standard Deviation of Discrete/Continuous Series = $\sqrt{\frac{\sum f(x - \bar{x})^2}{N}}$

4. Standard Deviation of Discrete/Continuous Series = $\sqrt{\frac{\sum fx^2}{N} - (\bar{x})^2}$

5. Coefficient of Variation = $\frac{SD}{AM} \times 100$

6. Variance = Square of Standard Deviation

7. Standard Deviation of first n natural numbers = $\sqrt{\frac{n^2 - 1}{12}}$

8. Standard Deviation of only two numbers a and b = $\frac{|a - b|}{2}$

9. If two variables x and y are related as $y = a + bx$, and SD_x is known, then
 $SD_y = |b| \times SD_x$.

10. Combined Standard Deviation = $\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}}$

Here, $d_1 = \bar{x}_1 - \bar{x}$; $d_2 = \bar{x}_2 - \bar{x}$

Topic 12 – Quartile Deviation

1. Quartile Deviation = $\frac{Q_3 - Q_1}{2}$

2. Coefficient of Quartile Deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$

3. If two variables x and y are related as $y = a + bx$, and QD_x is known, then $QD_y = |b| \times QD_x$.

Chapter 15 – Probability

1. $\text{Probability} = \frac{\text{No. of Favourable Cases/Events/Outcomes}}{\text{Total No. of Cases/Events/Outcomes}}$

2. When an experiment with total number of events a is repeated b number of times, the total number of outcomes is given by a^b .

3. $\text{Odds in Favour of Event } A = \frac{\text{Number of Favourable Outcomes}}{\text{Number of Unfavourable Outcomes}}$

4. $\text{Odds Against Event } A = \frac{\text{Number of Unfavourable Outcomes}}{\text{Number of Favourable Outcomes}}$

5. If two events A and B are not mutually exclusive, then,
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

6. Two events A and B are mutually exclusive, if $A \cap B = \phi$. Therefore, $P(A \cap B) = 0$, or $P(A \cup B) = P(A) + P(B)$.
7. If three events A , B , and C are not mutually exclusive, then $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$
8. Three events A , B , and C are mutually exclusive, if $P(A \cup B \cup C) = P(A) + P(B) + P(C)$.
9. Probability that only event A occurs: $P(A - B) = P(A) - P(A \cap B)$
10. Probability that only event B occurs: $P(B - A) = P(B) - P(A \cap B)$
11. $P(A \cap B) = P(A) \times P(B)$
12. Probability of event A given that event B has already occurred is given by $P(A/B)$:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

13. Probability of event B given that event A has already occurred is given by $P(B/A)$:

$$P(B/A) = \frac{P(B \cap A)}{P(A)}$$

14.
$$P(A'/B) = \frac{P(A' \cap B)}{P(B)} = \frac{P(B) - P(A \cap B)}{P(B)}$$

15.
$$P(A/B') = \frac{P(A \cap B')}{P(B')} = \frac{P(A) - P(A \cap B)}{1 - P(B)}$$

$$16. \quad P(A' \cap B') = \frac{P(A' \cap B')}{P(B')} = \frac{P(A \cup B)'}{1 - P(B)} = \frac{1 - P(A \cup B)}{1 - P(B)}$$

17. Probability that only event A or only event B occurs:

$$P(A) + P(B) - 2P(A \cap B)$$

18. Expected value (μ) of a random variable (x) is given by: $\mu = E(x) = \sum p_i x_i$

19. Expected value of (x^2) is given by: $E(x^2) = \sum [p_i (x_i^2)]$

20. Variance (σ^2) of a random variable (x) is given by:

$$V(x) = \sigma^2 = E(x - \mu)^2 = E(x^2) - \mu^2$$

21. If a and b are two constants related with two random variables x and y as $y = a + bx$, then the mean, i.e., the expected value of y is given by: $\mu_y = a + b\mu_x$.

22. If a and b are two constants related with two random variables x and y as $y = a + bx$, then the standard deviation of y is given by: $\sigma_y = |b| \times \sigma_x$.
23. Expectation of sum of two random variables is the sum of their expectations, i.e., $E(x + y) = E(x) + E(y)$, for any two random variables x and y .
24. Expectation of the product of two random variables is the product of the expectation of the two random variables, provided the two variables are independent, i.e., $E(x \times y) = E(x) \times E(y)$.

Chapter 16 – Theoretical Distributions

Topic 1 – Binomial Distributions

1. $P(x) = {}^nC_x p^x q^{n-x}$, for $x = 0, 1, 2, 3, \dots, n$
2. The mean of the binomial distribution is given by $\mu = np$.
3. A binominal distribution is symmetrical when $p = q$.
4. Mode of a Binomial Distribution is given by largest integer contained in $\mu_0 = (n+1)p$, OR $\mu_0 = (n+1)p$ and $[(n+1)p] - 1$

5. The variance of the binomial distribution is given by $\sigma^2 = npq$.
6. If $p = q = 0.5$, variance is the maximum, and is given by $\frac{n}{4}$.
7. Standard Deviation of a binomial distribution is given by $\sigma = \sqrt{npq}$.
8. Let x and y be two independent binomial distributions where x has the parameters n_1 and p , and y has the parameters n_2 and p . Then $(x + y)$ will be a binomial distribution with parameters $(n_1 + n_2)$ and p .

Topic 2 – Poisson Distribution

1. $P(x) = \frac{e^{-m} \times m^x}{x!}$, for $x = 0, 1, 2, 3, \dots, n$
2. The mean of Poisson distribution is given by m , i.e., $\mu = m = np$.
3. The variance of Poisson distribution is given by $\sigma^2 = m = np$.
4. The standard deviation of Poisson distribution is given by $\sigma = \sqrt{m} = \sqrt{np}$.
5. Mode of a Poisson Distribution is given by largest integer contained in m , OR m and $m - 1$
6. Let x and y be two independent poisson distributions where x has the parameter m_1 , and y has the parameter m_2 . Then $(x + y)$ will be a poisson distribution with parameter $(m_1 + m_2)$.

Topic 3 – Normal Distribution

$$1. P(x) = f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)}, \text{ for } -\infty < x < \infty$$

2. If $\mu = 0$, and $\sigma = 1$, the variable is known as the standard normal variate (or variable).

It is given by $z = \frac{x - \mu}{\sigma}$.

3. Relationship between MD, SD, and QD $\rightarrow 4SD = 5MD = 6QD$

4. Mean Deviation $= 0.8\sigma$.

5. Quartile Deviation $= 0.675\sigma$.

$$6. Q_1 = \mu - 0.675\sigma$$

$$7. Q_3 = \mu + 0.675\sigma$$

8. Median $- Q_1 = Q_3 - \text{Median}$.

9. Points of inflexion are given by $x = \mu - \sigma$ and $x = \mu + \sigma$

10. If there are two Independent Normal Distributions $x \sim N(\mu_1, \sigma_1^2)$ and $y \sim N(\mu_2, \sigma_2^2)$, then $z = x + y$ follows normal distribution with mean $(\mu_1 + \mu_2)$ and $SD = \sqrt{\sigma_1^2 + \sigma_2^2}$ respectively.

Chapter 17 – Correlation and Regression

Topic 1 – Karl Pearson's Product Moment Correlation Coefficient

$$1. r = r_{xy} = \frac{Cov(x, y)}{S_x \times S_y} = \frac{Cov(x, y)}{\sigma_x \times \sigma_y}$$

$$a. Cov(x, y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{n} = \frac{\sum xy}{n} - \bar{x} \cdot \bar{y}$$

$$b. S_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n}} = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$$

$$c. S_y = \sqrt{\frac{\sum (y - \bar{y})^2}{n}} = \sqrt{\frac{\sum y^2}{n} - \bar{y}^2}$$

$$2. r = \frac{n \sum xy - \sum x \times \sum y}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}}$$

3. Let there be two variables x and y . Let the correlation coefficient between them be r_{xy} . Now, if they are changed to another set of variables, say, u and v , then,
- $r_{uv} = r_{xy}$, if b and d have the same sign, or
- $r_{uv} = -r_{xy}$, if b and d have opposite signs.

Here,

$$b = \frac{-\text{Coefficient of } u}{\text{Coefficient of } x}, \text{ and } d = \frac{-\text{Coefficient of } v}{\text{Coefficient of } y}$$

$$4. r_{xy} = \frac{bd}{|b||d|} \cdot r_{uv}$$

Topic 2 – Spearman's Rank Correlation Coefficient

$$1. r_R = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

$$2. \text{ When ranks repeat, then } r_R = 1 - \frac{6 \left[\sum d^2 + \frac{\sum (t_j^3 - t_j)}{12} \right]}{n(n^2 - 1)}$$

Topic 3 – Coefficient of Concurrent Deviations

1. If $(2c - m) > 0$, then $r_c = \sqrt{\frac{(2c - m)}{m}}$

2. If $(2c - m) < 0$, then $r_c = -\sqrt{-\frac{(2c - m)}{m}}$

Topic 4 – Regression

1. if y depends on x , then the regression line of y on x is given by:

a. either $y = a + bx$, where,

i. $b = b_{yx} = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$, or, $b_{yx} = r \frac{\sigma_y}{\sigma_x}$, or, $b_{yx} = \frac{Cov(x, y)}{(\sigma_x)^2}$, and

ii. $a = a_{yx} = \bar{y} - (\bar{x} \times b_{yx})$

b_{yx} is known as the regression coefficient.

b. or, $(y - \bar{y}) = b_{yx}(x - \bar{x})$

2. if x depends on y , then the regression line of x on y is given by:

a. either $x = a + by$, where,

$$\text{i. } b = b_{xy} = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum y^2 - (\sum y)^2}, \text{ or, } b_{xy} = r \frac{\sigma_x}{\sigma_y}, \text{ or, } b_{xy} = \frac{\text{Cov}(x, y)}{(\sigma_y)^2}, \text{ and}$$

$$\text{ii. } a = a_{xy} = \bar{x} - (\bar{y} \times b_{xy})$$

b_{xy} is known as the regression coefficient.

$$\text{or, } (x - \bar{x}) = b_{xy} (y - \bar{y})$$

$$3. b_{vu} = \frac{\text{Scale of } v}{\text{Scale of } u} \times b_{yx}$$

$$4. r = \sqrt{b_{yx} \times b_{xy}}$$

Topic 5 – Probable Error and Standard Error

1. Probable Error (P.E.) is given by

a. $P.E. = 0.674 \times \frac{1-r^2}{\sqrt{N}}$, or

b. $P.E. = 0.6745 \times \frac{1-r^2}{\sqrt{N}}$, or

c. $P.E. = 0.675 \times \frac{1-r^2}{\sqrt{N}}$

2. Limits of the correlation coefficient of the population is given by $p = r \pm P.E.$

3. Standard Error (S.E.) is given by $S.E. = \frac{1-r^2}{\sqrt{N}}$

$$4. PE = \frac{2}{3} SE$$

Topic 6 – Coefficient of Determination and Non-Determination

1. Coefficient of Determination (Also known as “Percentage of Variation Accounted for”) $(r^2) = \frac{\text{Explained Variance}}{\text{Total Variance}}$
2. Coefficient of Non-Determination (Also known as “Percentage of Variation Unaccounted for”) $= 1 - r^2$

Chapter 18 – Index Numbers

Topic 1 – Price Relative

1. Price Relative = $\frac{P_n}{P_0} \times 100$.

2. Simple Aggregative Price Index = $\frac{\sum P_n}{\sum P_0} \times 100$

3. Simple Average of Price Relatives = $\frac{\sum \left(\frac{P_1}{P_0} \times 100 \right)}{N}$

$$4. \text{Weighted Average General Index} = \frac{\text{Sum of Products}}{\text{Sum of Weights}}$$

Topic 2 – Weighted Aggregative Index

$$1. \text{ Laspeyres Index} = \frac{\sum P_n Q_0}{\sum P_0 Q_0} \times 100$$

$$2. \text{ Passche's Index} = \frac{\sum P_n Q_n}{\sum P_0 Q_n} \times 100$$

$$3. \text{ Marshall-Edgeworth Index} = \frac{\sum P_n (Q_0 + Q_n)}{\sum P_0 (Q_0 + Q_n)} \times 100$$

$$4. \text{ Bowley's Index} = \frac{\text{Laspeyres' + Paasche's}}{2}$$

$$5. \text{ Fisher's Index} = \sqrt{\frac{\sum P_n Q_0}{\sum P_0 Q_0} \times \frac{\sum P_n Q_n}{\sum P_0 Q_n}} \times 100$$

CA NISHANT KUMAR

Topic 3 – Weighted Average of Price Relatives

$$1. \text{ Weighted Average} = \frac{\sum \left[\frac{P_n}{P_0} \times (P_0 Q_0) \right]}{\sum P_0 Q_0} \times 100$$

Topic 4 – The Chain Index Numbers

$$1. \text{Chain Index} = \frac{\text{Link Relative of the Current Year} \times \text{Chain Index of the Previous Year}}{100}$$

Topic 5 – Quantity Index Numbers

1. Simple Aggregate of Quantities = $\frac{\sum Q_n}{\sum Q_0} \times 100$

2. Simple Average of Quantity Relatives = $\frac{\frac{\sum Q_n}{\sum Q_0}}{n} \times 100$

3. Laspeyre's Index = $\frac{\sum Q_n P_0}{\sum Q_0 P_0} \times 100$

6. Paasche's Index = $\frac{\sum Q_n P_n}{\sum Q_0 P_n} \times 100$

$$7. \text{ Fisher's Index} = \sqrt{\frac{\sum Q_n P_0}{\sum Q_0 P_0} \times \frac{\sum Q_n P_n}{\sum Q_0 P_n}} \times 100$$

$$8. \text{ Base-year weighted average of quantity relatives} = \frac{\sum \left[\frac{Q_n}{Q_0} \times (P_0 Q_0) \right]}{\sum P_0 Q_0} \times 100$$

Topic 6 – Value Index Numbers

$$\text{Value Index} = \frac{\sum V_n}{\sum V_0} = \frac{\sum P_n Q_n}{\sum P_0 Q_0}$$

Topic 7 – Deflating Time Series Using Index Numbers

$$1. \text{ Deflated Value} = \frac{\text{Current Value}}{\text{Price Index of the Current Year}}$$

$$2. \text{ Current Value} = \frac{\text{Base Price } (P_0)}{\text{Current Price } (P_n)}$$

$$3. \text{ Real Wages} = \frac{\text{Actual Wages}}{\text{Cost of Living Index}} \times 100$$

Topic 8 – Shifting and Splicing of Index Numbers

$$1. \text{ Shifted Price Index} = \frac{\text{Original Price Index}}{\text{Price Index of the Year on which it has to be shifted}} \times 100$$

Topic 9 – Test of Adequacy

1. Time Reversal Test $P_{01} \times P_{10} = 1$
2. Factor Reversal Test $P_{01} \times Q_{01} = V_{01}$
3. Circular Test $P_{01} \times P_{12} \times P_{20} = 1$